On the development of QSPR models for regulatory frameworks : The heat of decomposition of nitroaromatics as a test case

Guillaume Fayet, Patricia Rotureau, Carlo Adamo

On the development of QSPR models for regulatory frameworks: the heat of decomposition of nitroaromatics as a test case

Guillaume Fayet[a,]*, Patricia Rotureau[a], Carlo Adamo[b]

[a] INERIS, Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte, France

[b] Laboratoire d'Electrochimie, Chimie des Interfaces et Modélisation pour l'Energie, CNRS UMR-7575, Chimie ParisTech, 11 rue P. et M. Curie, 75231 Paris Cedex 05, France

*Corresponding author: guillaume.fayet@ineris.fr; tel: +33(0)344618126; fax: +33(0)344556565

Abstract

Many regulatory frameworks, e.g. related to the Transport of Dangerous Goods, the Registration, Evaluation, Authorisation and restriction of chemicals (REACH) or the Classification, Labelling and Packaging of substances and mixtures (CLP), require the characterization of the hazards of chemicals, which could be complex. In particular, the REACH regulation involves an extensive quantity of works, to gather toxicological, eco-toxicological and physico-chemical properties for a large number of compounds. So, the full characterization by experimental way is time-consuming and cost-expensive. Alternative methods are therefore encouraged to complement experimental tests. The Quantitative Structure-Property Relationships (QSPR) approach is one of the recommended methods, provided that they are developed within the rigorous guidelines proposed by the Organization for Economic Co-operation and Development (OECD). In this context, a series of nitroaromatic compounds has been analyzed to achieve new QSPR models for the prediction of their heat of decomposition respecting the requirements for application in regulatory frameworks.

Three multilinear models were obtained upon the set of descriptors considered for their development (constitutional, topological or both) that do not need any preliminary time expensive quantum chemical calculations. They were tested by internal and external validation tests. Good performances for the two ones including constitutional descriptors were obtained in particular in terms of predictive power in a well defined applicability domain (R^2_{IN}=0.81-0.87). They are easier to apply than our previous quantum chemical based model, since they do not need any preliminary calculations.

Keywords: Quantitative Structure-Property Relationships; heat of decomposition; nitroaromatic compounds; constitutional and topological descriptors

Highlights:

- QSPR models were developed for the heat of decomposition of nitroaromatic compounds.
- Two accurate MLR models were exhibited based on simple constitutional (and topological) descriptors.
- Performances were evaluated by a series of internal and external validations.
- The new QSPR models satisfied all OCDE principles of validation for regulatory use.

1 Introduction

To characterize the intrinsic hazard of chemical compounds, the Recommendations on the Transport of Dangerous Goods (UN, 2011) listed a series of tests. Among the involved physico-chemical properties, heats of decomposition are considered as a pre-selection criterion to identify substances that could present explosive properties. Measured by calorimetric analyses, the heat of decomposition evaluates the amount of energy released during the decomposition of chemicals. Among available methods, differential scanning calorimetry (DSC) (Chervin & Bodman, 2003; Grewer, 1994; Jones & Augsten, 1996; Yoshida, 1987) providing heats of decomposition with uncertainties of measurement of about 5–10% (Ando, Fujimoto, & Morisaki, 1991) is a typical example of an experimental screening test.

Such tests are also considered in the new European REACH (EC 1907/2006) and CLP (EC 1272/2008) regulations. This new regulatory framework implies the systematic characterization of a tremendous number of substances, since more than 143 000 were submitted in the pre-registration phase of REACH between June and December 2008 (EChA, 2012). Unfortunately, such volume of work is incompatible with the imposed calendar, i.e. until 2018 for existing substances (depending on quantities of chemicals produced or imported). So, the use of methods, alternative to experimental testing, such as quantitative structure property relationships (QSPR), was explicitly recommended to obtain information data (physico-chemical, toxicological and eco-toxicological) required by REACH.

Indeed, QSPR models represent powerful tools already successfully used for biological (Winkler, 2002), toxicological (Cronin & Worth, 2008; Netzeva, Pavan, & Worth, 2008), pharmaceutical (Grover, Singh, Bakshi, & Singh, 2000a, b) and physico-chemical applications (Dearden & Worth, 2007; Katritzky et al., 2010). Five principles have even been proposed by the Organization for Economic Co-operation and Development (OECD) for their validation in the context of regulatory uses (OECD, 2007). At first, the endpoints of models have to be fully defined, i.e. including the description of experimental protocols. Secondly, algorithm must be transparent, so that the model equations (or structures) and all the related computational parameters must be clearly defined. Then, domains of applicability must be defined so to determine on which systems accurate predictions are expected or not. Performances have to be evaluated on a series of validation tests, including the characterization of predictivity on an external set of compounds. Finally, the fifth principle recommends, when possible, a mechanistic interpretation of models to link their parameters to the subjacent mechanisms involved in the studied properties. These principles have already been taken into account for the development of various QSPR models in the last years for toxicological endpoints (Benigni, Bossa, Netzeva, & Worth, 2007; Gramatica, 2007) as for physico-chemical properties (Fayet, Del Rio, Rotureau, Joubert, & Adamo, 2011; Fayet, Rotureau, Joubert, & Adamo, 2011; Öberg & Liu, 2011; Papa, Kovarich, & Gramatica, 2009; Prana, Fayet, Rotureau, & Adamo, in press). Moreover, the applicability of such models in an industrial context has been demonstrated (Patlewicz, Chen, & Bellin, 2011).

Concerning the heat of decomposition of nitroaromatic compounds, only few QSPR studies have been carried out. First attempts used limited datasets that did not allow any external validation. Saraf proposed a correlation with the number of nitro groups (Saraf, Rogers, & Mannan, 2003) with a fitting error of 8% for a very limited number of compounds (19 nitrobenzene derivatives). In previous works, a series of preliminary multilinear models were derived from a data set of 22 molecules with

high correlations (up to R^2=0.98) by introducing quantum chemical descriptors (Fayet, Joubert, Rotureau, & Adamo, 2009a; Fayet, Rotureau, Joubert, & Adamo, 2009, 2010). Unfortunately, the size of these datasets did not allow any external validation. In more recent works (Fayet, Del Rio, Rotureau, Joubert, & Adamo, 2011; Fayet, Rotureau, Joubert, & Adamo, 2011), we obtained very robust models, following the OECD principles, by analyzing a more extended data set of 77 nitrobenzene derivatives. An accurate model for the whole diversity of structures included into this dataset was obtained through a qualitative decision tree with high predictivity (82% of correct predictions obtained on a validation set of 22 molecules) (Fayet, Del Rio, Rotureau, Joubert, & Adamo, 2011). Another reliable model (Fayet, Rotureau, Joubert, & Adamo, 2011), based on a quantitative multilinear regression with a predictivity in its applicability domain of R^2_{IN}=0.86, was issued from a data set restricted to the compounds presenting no substituent in ortho position to the nitro group considering the fact that the presence of such ortho substituent possibly involves particular decomposition mechanisms. However, these two models require a preliminary determination of quantum chemical properties which could be time-consuming and not necessarily straightforward.

Starting from these “proofs of principles” on the possibility to develop robust QSPR models respecting OECD principles, new multilinear models were developed based on the same data set, in particular for non-ortho substituted nitroaromatics, by considering only simple constitutional and topological descriptors. Based on these descriptors, models will be easier to apply by industrials in the context of REACH for example, since they do not need any prior quantum chemical calculations.

2 Materials and methods

2.1 Experimental data set

Heats of decomposition (-ΔH) of 42 nitroaromatic compounds have been used to derive high predictive models. Experimental data represent a critical point of the QSPR analysis, since they have to be obtained following a single protocol in order to satisfy the first principle of the OECD guidelines (OECD, 2007). Indeed, this ensures the compatibility between data and then it reduces uncertainties that could propagate in the model during the fitting procedure. For this reason, all experimental data were extracted from a single reference (Ando, Fujimoto, & Morisaki, 1991). The heats of decomposition were measured using a pressure differential scanning calorimetry (DSC) apparatus, on 1–2 mg samples in aluminum cells with pin-hole, with a heat rate of 10 K/min.

It has to be noticed that the studied compounds (presented in table 1) consisted in nitrobenzene derivatives substituted by a variety of groups (e.g. nitro, amino or halogens), with the particularity to present no substituent in ortho position to a nitro group. Indeed, ortho substituted compounds potentially undergo to specific decomposition mechanism involving the interaction of the nitro group with the adjacent substituent as evidenced on nitrotoluene derivatives (Fayet, Joubert, Rotureau, & Adamo, 2009b; Fayet, Rotureau, Joubert, & Adamo, 2011).

To evaluate the predictivity of the developed models, this data set was divided into two parts. A training set of 31 molecules was used for the development of the model and a validation set of 11 compounds was used to compute an external validation. The partition between sets defined in previous work (Fayet, Rotureau, Joubert, & Adamo, 2011) was kept as it ensured the same

distribution in property values in both training and validation sets (as shown in Figure 1). Moreover, no bias in the diversity of chemical structures was observed in each set.

2.2 Molecular descriptors

The molecular structures of the selected nitroaromatic compounds were characterized by a series of 88 molecular descriptors that can be extracted from a simple 2D structure. Constitutional descriptors are the simplest descriptors that reflect the molecular composition. They count the number of specific atoms or bonds in molecules (e.g., number of O atoms and number of single bonds). Topological descriptors, like the Wiener index, are based on atomic connectivity tables and provide information about the size and shape of molecules. All these descriptors do not require quantum chemical computations and have been evaluated using the Codessa software (Codessa, 2002). More information is available in the books of Karelson (1996) and Todeschini (2000). In addition, external descriptors that do not need any expensive calculation times were added, like the occurrence and count of specific molecular groups identified in the molecules of the database (e.g. number of nitro groups).

2.3 QSPR modeling

In this paper, multilinear regressions were derived. Such models are based on the following formula:

$$Y = a_0 + \sum_i a_i X_i \qquad \text{(Eq. 1)}$$

where Y is the calculated property, X_i are the molecular descriptors and a_i the regression constants.

To be accurate, models have to be constituted by an optimized set of descriptors. Indeed, a too large number of descriptors involve large errors in prediction due to the inclusion of parameters that are not really related to the property and to inter-correlated descriptors that represent redundant pieces of information. So, to achieve reliable QSPR models, a parameter selection, using the Best Multi Linear Regression technique, as implemented in Codessa software, was realized among the 88 calculated descriptors.

This stepwise approach , already successfully used in previous works (Fayet, Rotureau, Joubert, & Adamo, 2010, 2011), started with constructing two-parameter models based on non-intercorrelated descriptors (with R^2 between descriptors lower than 0.1) and then it built higher rank models by adding new non-intercorrelated descriptors (i.e. with R^2 lower than 0.6 with each of the previous ones). By this way, this method guarantees that two intercorrelated descriptors are not selected in the same model. Finally, the algorithm gave, at each rank (i.e. for each number of descriptors), the model presenting the higher correlation with the studied property.

The final model was chosen among these regressions as the one representing the best compromise between the correlation refinement and the number of descriptors. The pertinence of each descriptor in the model was also checked based on a student t-test at a 95% level of confidence (presented in Supporting Information, Tables S1-S3).

2.4 Internal and external validations

To investigate the performances of the developed models, a series of internal and external validation tests were performed (Tropsha, 2010; Witten & Frank, 2005). The goodness of fit was evaluated for the molecules of the training set by the coefficient of determination (R^2) and the relative mean absolute error (MAE_{TR}(%)).

Leave-one-out (LOO) and leave-many-out (LMO) cross validations were computed, to estimate the robustness of the model, i.e. the dependence of the fitting of the model to any molecule(s) of the training set, via the Q^2_{LOO}, Q^2_{10CV} and Q^2_{5CV} coefficients (for LOO, 10-fold and 5-fold cross validations, respectively). These coefficients were expected to be stable upon partition size and close to R^2.

The models were also evaluated against chance correlation by Y-randomization (Rücker, Rücker, & Meringer, 2007). Property values were randomized within the training set by 500 successive iterations. From each new randomized data set, a new model was computed again, with performances expected to be low. Finally, the average value and the standard deviation in R^2 coefficients for the randomized models (denoted R^2_{YS} and SD_{YS}, respectively) were calculated, to check that the original model was strongly more performant than the randomized ones.

Then, the predictive powers of the models were estimated for the molecules of the validation set by the coefficient of determination (R^2_{EXT}), the relative mean absolute error (MAE_{EXT}(%)) and the Q^2_{EXT} coefficient proposed by OECD guidelines (OECD, 2007).

2.5 Applicability Domain

The applicability domain (AD) of each model, i.e. the domain in which predictions can be considered as accurate, was defined by the molecules of the training set. It was built, for each descriptor, by the range of values represented among the molecules of the training set. The AD ranges for each descriptor in each model are available in Supporting Information (Table S4).

So, for new predictions, future users will simply have to preliminary check if descriptor values are in the intervals defining the AD to know if the model is applicable for the molecules they want to consider.

To evaluate the real predictive capability of the models in the context of future predictions, the predictive performances of all models were finally re-evaluated taking into account ADs (noticed R^2_{IN}, MAE_{IN} (%) and Q^2_{IN}, respectively).

3 Results

3.1 Model based on constitutional descriptors

A first model was developed based only on 50 constitutional descriptors. The BMLR procedure proposed equations including up to 16 descriptors and the best compromise between the correlation and the number of descriptors was obtained for a four-parameter model:

$$-\Delta H = -594.5 + 2381.6\, n_{db,rel} + 306.5\, n_{NO2} - 791.4\, n_{O,rel} + 83.4\, n_{conj} \qquad \text{(Eq. 2)}$$

where $n_{db,rel}$ and $n_{O,rel}$ are the relative numbers of double bonds and oxygen atoms, respectively, and n_{NO2} and n_{conj} are the numbers of nitro groups and conjugated bonds, respectively. It has to be noticed that nitro groups contain two conjugated bonds.

In this equation, all descriptors presented the same importance in the regression with close absolute values of t-test (from 2.2 to 5, in table S1). The presence of n_{NO2} should be particularly noticed. Indeed, already included in previous models addressing the prediction of the same property (Fayet, Del Rio, Rotureau, Joubert, & Adamo, 2011; Fayet, Rotureau, Joubert, & Adamo, 2011), this descriptor is pertinent from a chemical point of view since the energy released during decomposition is linked to the loss of nitro groups (Brill & James, 1993).

From a statistical point of view, this model performed quite well with R^2=0.84 (see table 2 and Figure 2). Internal validations were also satisfactory. LOO and LMO cross-validation coefficients (Q^2) ranged between 0.78 and 0.80 and Y-randomization ensured against a chance correlation with an average correlation coefficient R^2_{YS}=0.13 and a low standard deviation (SD_{YS}=0.08) over the 500 randomization iterations (as shown in Figure 3). Finally, the predictive power was remarkable with R^2_{EXT}=0.81 and Q^2_{EXT}=0.81, in particular in its AD since no molecule of the validation set was out of it (so, R^2_{IN}=0.81 and Q^2_{IN}=0.81).

3.2 Model based on topological descriptors

In a second step, 38 topological descriptors were considered. After the BMLR analysis, the following three-parameter regression in Eq. 3 was selected as the one presenting the most important correlation regarding its level of parameterization.

$$-\Delta H = -1385.4 + 142.0\ {}^{0}\chi + 953.4\ {}^{1}IC_{avg} - 31.8\ {}^{1}IC \quad \text{(Eq. 3)}$$

where ${}^{0}\chi$ is the Randic index (order 0), ${}^{1}IC_{avg}$ is the average information content (order 1) and ${}^{1}IC$ is the information content (order 1).

The interpretation of the topological descriptors was more difficult since they mainly represent the shape of molecules. In the cases of biological effect, they are very likely understood as representing a molecular feature that interacts with biological receptors. In the case of impact sensitivity, subjacent mechanism is very different and does not issue from such steric interaction. If these descriptors also characterize the size of molecules, which is globally connected to the quantity of energy available for the decomposition, such direct interpretation was not evidenced since none of these descriptors was related to the chemical mechanisms involved in the decomposition of nitro compounds.

The performances of this model were less good than the previous one (Eq. 2), in terms of correlation (R^2=0.78) and robustness (Q^2_{LOO}=0.71). Nevertheless, the Y-randomization test exhibited no chance correlation and predictions, once considering the applicability domain of the model appeared satisfactory (R^2_{IN}=0.82).

It has to be noticed that R^2_{EXT} value was lower (0.46) due to the large error in the prediction for the nitrobenzene molecule (more than 300 kJ/mol in error) which was not included into the AD of the model. This shows the importance of taking into account the applicability domain in the evaluation of the predictive power.

The slight increase of values between R^2 and R^2_{IN} (0.04) could be considered in the present case as due to the use of small sets of data for both the training and the validation sets. However, both correlation and predictivity are significant.

3.3 Model based on both constitutional and topological descriptors

In a last step, the whole set of 88 constitutional and topological descriptors was used to derive the final four-parameter model, which associated the best correlation regarding its number of descriptors among the regressions issued from BMLR analysis:

$$-\Delta H = 153.5 + 386.4\ n_{NO2} - 78.9\ n_{CH3} + 131.2\ {}^{1}\chi^{v} - 327.7\ {}^{0}IC_{avg} \quad \text{(Eq. 4)}$$

where n_{CH3} is the number of methyl groups, ${}^{1}\chi^{v}$ is the Kier and Hall index (order 1), ${}^{0}IC_{avg}$ is the average information content (order 0).

If n_{CH3}, ${}^{1}\chi^{v}$ and ${}^{0}IC_{avg}$ are not directly interpretable in terms of chemical mechanism, the number of nitro groups was satisfactory included (like for the constitutional-based model) since it is in agreement with the general statements recognized on the decomposition process of nitro compounds. Besides this descriptor was the most significant in the model when considering its t-test value (9.30).

The correlation of this model was also high with R^2=0.85 and stable among the series of computed cross validation (Q^2_{LOO}=Q^2_{5CV}=0.76, Q^2_{10CV}=0.74). It did not issue from a chance correlation as demonstrated by the 500-iteration Y-randomization test (R^2_{YS}=0.13, SD_{YS}=0.08). Finally, the predictive power calculated for the 11 molecules of the validation set was also good (with R^2_{EXT}=0.81). When considering the only compounds that were included in the AD of the model, R^2_{IN} value (0.87) was slightly higher than R^2 but, regarding experimental uncertainties (5-10%) and the size of the dataset, the difference between R^2_{IN} and R^2 (0.02) was not significant. Moreover, both R^2 and R^2_{IN} values were particularly remarkable.

4 Discussion

The three models developed in this study respect the OECD principles of validation for regulatory uses. Indeed, these models were defined starting from heats of decomposition obtained from an unique original reference to ensure that data were obtained using a single protocol. The algorithms of the models are simple and entirely defined on constitutional and topological descriptors. They are applicable for nitrobenzene derivatives that present no substituent in ortho position to the nitro group in an AD simply defined on the values of the descriptors included in the models. The detailed ranges of values for each descriptor in each model are available in supporting information (Table S4). Performances were revealed from internal and external validation tests: fitting evaluation for the molecules of the training set, LOO and LMO cross-validations, Y-randomization, predictions on an external validation set.

From the three developed models, the ones presenting the best performances are those integrating constitutional descriptors (Eqs. 2 and 4). Indeed, they do not only propose significant predictivity in their applicability domains (R^2_{IN}=0.81 and 0.87, for Eqs. 2 and 4 respectively) but they are also satisfactory in terms of fitting and robustness (with R^2=0.84-0.85 and Q^2_{LOO}=0.76-0.79). Moreover, these two models include the number of nitro groups that is directly related to the energy released during decomposition. The topological model (Eq. 3) presents globally inferior performances in particular in terms of correlation and robustness (R^2=0.78 and Q^2_{LOO}=0.71). Moreover, no straightforward interpretation is evidenced for any of the constituting topological descriptors.

It has to be noticed that the n_{NO2} descriptor introduced in models of Eqs. 2 and 4 was already included in the decision tree developed in our previous work (Fayet, Del Rio, Rotureau, Joubert, & Adamo, 2011; Fayet, Rotureau, Joubert, & Adamo, 2011) for the prediction of the heat of decomposition of nitroaromatic compounds. Moreover, descriptors influenced by the presence of NO_2 groups were also selected in the MLR model obtained for the only non-ortho substituted nitroaromatic compounds (Fayet, Rotureau, Joubert, & Adamo, 2011, which represent the only previous quantitative model to have been externally validated to predict the heat of decomposition of nitroaromatic compounds.

$$-\Delta H = 0.8\ G - 3.8\ WPSA1 - 4255.1\ Q_{max} + 26.8\ RPCS - 251.2 \qquad \text{(Eq. 5)}$$

where G is the gravitation index, WPSA1 the weighted positive surface area, Q_{max} the maximal partial charge in the molecule and RPCS the relative positive charged surface area. This model was well correlated (R^2=0.90), robust (Q^2_{LOO}=0.86), and presented remarkable predictive power for an external validation set (R^2_{EXT}=0.84), in particular in its applicability domain (R^2_{IN}=0.86).

Compared to this model including quantum chemical descriptors (Fayet, Rotureau, Joubert, & Adamo, 2011), the two new models (Eq. 2 and 4) are slightly less fitted, for the molecules of the training set (R^2=0.84-0.85). Moreover, a slight decrease in accuracy is observed from the predictions computed for the molecules of the validation set for the first model (R^2_{IN}=0.81 in this paper vs. R^2_{IN}=0.86 in our previous work). Nevertheless, the last model (Eq. 4), including both constitutional and topological descriptors, presents similar predictive ability (R^2_{IN}=0.87).

Globally, the performances of these two new models remain very interesting in view of predictions, with nearly similar capabilities and less computer times than our previous model, which needed prior quantum chemical calculations. In particular, the constitutional-based model (Eq. 2) is very simple with high performances and represents the best compromise between performances and practical complexity for final users to obtain predicted data.

5 Conclusion

In this paper, new multilinear QSPR models were developed to predict the heats of decomposition of nitroaromatic compounds. In particular, the target molecules were nitrobenzene derivatives not substituted in ortho position to the nitro group. These models were derived from a series of constitutional and topological descriptors with the aim to achieve reliable predictions without any time expensive calculations. Three models were computed according to the OECD principles for validation for regulatory use from a dataset of 42 compounds using constitutional and/or topological descriptors. As a consequence, they can be used in a regulatory framework like REACH.

High performances were exhibited in terms of correlation, robustness (including leave-many-out cross validations), absence of chance correlation (by Y-randomization) and predictive power for the two models including constitutional descriptors (Eqs. 2 and 4). Moreover, these two models were based on the number of nitro groups that is recognized to be linked to the energy released during the decomposition process. Finally, they present performances close to the ones of the previous quantum chemical model and the advantage to be easier to apply (without any time expensive preliminary calculations).

The model based only on constitutional descriptors was particularly interesting since it was very easy to use for prediction to any user without needing any complex calculations.

6 References

Ando, T., Fujimoto, Y., & Morisaki, S. (1991). Analysis of differential scanning calorimetric data for reactive chemicals. Journal of Hazardous Materials, 28, 251-280.

Benigni, R., Bossa, C., Netzeva, T., & Worth, A. (2007). Collection and Evaluation of (Q)SAR Models for Mutagenicity and Carcinogenicity: European Commission, Joint Research Centre.

Brill, T. B., & James, K. J. (1993). Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives. Chemical Review, 93, 2667-2692.

Chervin, S., & Bodman, G. I. (2003). Method for estimating decomposition characteristics of energetic chemicals. Process Safety Progress, 22, 241-243.

Codessa. (2002). University of Florida.

Cronin, M. T. D., & Worth, A. P. (2008). (Q)SARs for Predicting Effects Relating to Reproductive Toxicity. QSAR & Combinatorial Science, 27, 91-100.

Dearden, J., & Worth, A. (2007). In Silico Prediction of Physicochemical Properties: European Commission, Joint Research Centre.

EC (European Commission) (2006). Regulation (EC) N° 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).

EC (European Commission) (2008). Regulation (EC) N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006.

EChA. (2012) Pre-registered substances. Accessed 20/02/2012, from http://echa.europa.eu/web/guest/information-on-chemicals/pre-registered-substances

Fayet, G., Del Rio, A., Rotureau, P., Joubert, L., & Adamo, C. (2011). Predicting the thermal stability of nitroaromatic Compounds using Chemoinformatic Tools. Molecular Informatics, 30, 623-634.

Fayet, G., Joubert, L., Rotureau, P., & Adamo, C. (2009a). On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility. Chemical Physics Letter, 467, 407-411.

Fayet, G., Joubert, L., Rotureau, P., & Adamo, C. (2009b). A theoretical study of the decomposition mechanisms on substituted ortho-nitrotoluenes. Journal of Physical Chemistry A, 113, 13621-13627.

Fayet, G., Rotureau, P., Joubert, L., & Adamo, C. (2009). On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations. Journal of Hazardous Materials, 171, 845-850.

Fayet, G., Rotureau, P., Joubert, L., & Adamo, C. (2010). QSPR Modeling of Thermal Stability of Nitroaromatic Compounds: DFT vs. AM1 Calculated Descriptors. Journal of Molecular Modelling, 16, 805-812.

Fayet, G., Rotureau, P., Joubert, L., & Adamo, C. (2011). Development of a QSPR model for predicting thermal stabilities of nitroaromatic compounds taking into account their decomposition mechanisms. Journal of Molecular Modelling, 17, 2443-2453.

Gramatica, P. (2007). Principles of QSAR models validation: internal and external. QSAR & Combinatorial Science, 26, 694-701.

Grewer, T. (1994). Thermal Hazards of Chemical Reactions. Amsterdam: Elsevier.

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000a). Quantitative structure-property relationships in pharmaceutical research - Part 1. Pharmaceutical Science & Technology Today, 3, 28-35.

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000b). Quantitative structure-property relationships in pharmaceutical research - Part 2. Pharmaceutical Science & Technology Today, 3, 50-57.

Jones, D. E. G., & Augsten, R. A. (1996). Evaluation of systems for use in DSC measurements on energetic materials. Thermochimica Acta, 286, 355-373.

Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-Chemical Descriptors in QSAR/QSPR Studies. Chemical Reviews, 96, 1027-1044.

Katritzky, A. R., Kuanar, M., Slavov, S., Hall, C. D., Karelson, M., Kahn, I., et al. (2010). Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. Chemical Reviews, 110, 5714-5789.

Netzeva, T. I., Pavan, M., & Worth, A. P. (2008). Review of (Quantitative) Structure–Activity Relationships for Acute Aquatic Toxicity. QSAR & Combinatorial Science, 27, 77-90.

Öberg, T., & Liu, T. (2011). Extension of a prediction model to estimate vapor pressures of perfluorinated compounds (PFCs). Chemometrics and Intelligent Laboratory Systems, 107, 59-64.

OECD (Organisation for Economic Co-operation and Development) (2007). Guidance Document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models (No. ENV/JM/MONO(2007)2).

Papa, E., Kovarich, S., & Gramatica, P. (2009). Development, Validation and Inspection of the Applicability Domain of QSPR Models for Physicochemical Properties of Polybrominated Diphenyl Ethers. QSAR & Combinatorial Science, 28, 790-796.

Patlewicz, G., Chen, M. W., & Bellin, C. A. (2011). Non-testing approaches under REACH - help or hindrance? Perspectives from a practitioner within industry. SAR and QSAR in Environmental Research, 22, 67-88.

Prana, V., Fayet, G., Rotureau, P., & Adamo, C. (in press). Predictive QSPR models for impact sensitivity of nitroaliphatic compounds. Journal of Hazardous Materials.

Rücker, C., Rücker, G., & Meringer, M. (2007). y-Randomization and Its Variants in QSPR/QSAR. Journal of Chemical Information and Modeling, 47, 2345-2357.

Saraf, S. R., Rogers, W. J., & Mannan, M. S. (2003). Prediction of reactive hazards based on molecular structure. Journal of Hazardous Materials, 98, 15-29.

Todeschini, R., & Consonni, V. (2000). Handbook of Molecular Descriptors. Weinheim: Wiley.

Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics, 29, 476-488.

UN (United Nations) (2011). Recommendations on the Transport of Dangerous Goods: Manual of Tests and Criteria. Fifth revised edition, ST/SG/AC.10/11/Rev.5 ed.

Winkler, D. A. (2002). The role of quantitative structure - activity relationships (QSAR) in biomolecular discovery. Briefings in Bioinformatics, 3, 73-86.

Witten, I. H., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. San Francisco: Morgan Kaufmann Publishers.

Yoshida, T. (1987). Safety of Reactive Chemicals (Vol. 1). Amsterdam: Elsevier.

Table 1 – Experimental and predicted heats of decomposition (in kJ/mol) of nitroaromatic compounds from new QSPR models (Eqs. 2-4) compared to the previous one (Eq. 5 (Fayet, Rotureau, Joubert, & Adamo, 2011))

ID	molecules	exp[b]	calculated			
			Eq. 2	Eq. 3	Eq. 4	Eq. 5
		training set				
1	2,6-dichloro-4-nitroaniline	264	280	325	315	284
2	2-amino-5-nitrophenol	153	240	189	248	239
3	3,5-dinitrobenzoic acid	674	686	656	717	679
4	3,5-dinitrobenzyl chloride	711	686	569	708	673
5	3-nitroacetoanilide	369	350	350	361	394
6	3-nitroaniline	350	280	300	301	317
7	3-nitroanisole	243	247	271	247	288
8	3-nitrocinnamic acid	414	425	488	419	417
9	3-nitrophenol	283	221	260	241	227
10	3-nitrotoluene	149	286	242	220	212
11	4-nitro-2-toluidine	306	296	185	228	315
12	4-nitroacetoanilide	387	350	350	361	372
13	4-nitroacetophenone	291	380	331	349	343
14	4-nitrobenzaldehyde	421	380	390	352	394
15	4-nitrobenzamide	319	331	336	350	321
16	4-nitrobenzhydrazide	362	344	374	370	335
17	4-nitrobenzyl alcohol	292	247	213	254	272
18	4-nitrophenol	232	221	260	241	235
19	4-nitrotoluene	213	286	242	220	192
20	2-amino-4-nitroanisole	375	261	234	260	325
21	2-amino-4-nitrophenol	130	240	189	248	173
22	3,5-dinitrobenzonitrile	654	667	659	614	698
23	3-nitrobenzoic acid	289	277	350	300	372
24	3-nitrobenzoic acid methylester	256	305	374	302	277
25	3-nitrophenylacetic acid	358	340	331	311	347
26	4-nitroaniline	347	280	300	301	308
27	4-nitrobenzoic acid methylester	302	305	374	302	264
28	4-nitrobenzoyl chloride	408	380	413	343	303
29	4-nitrobenzyl chloride	337	286	275	305	333
30	4-nitrophenetole	270	266	293	324	249
31	4-nitrophenylhydrazine	277	291	313	324	279
		validation set				
32	3,5-dinitrobenzamide	736	738	634	770	687
33	3-nitroacetophenone	276	380	331	349	364
34	3-nitrobenzaldehyde	373	380	390	352	389
35	3-nitrobenzamide	311	331	336	350	334
36	3-nitrobenzhydrazide	430	344	374	370	344
37	3-nitrobenzyl alcohol	325	247	213	254	258
38	4-nitroanisole	248	247	271	247	283

39	4-nitrobenzoic acid	284	277	350	300	332
40	4-nitrocinnamic acid	506	425	488	419	414
41	4-nitrophenylacetic acid	265	340	331	311	341[a]
42	nitrobenzene	161	266	481[a]	302[a]	202
		MAE_{TR}(%)	17	18	16	12
		MAE_{IN} (%)	18	15	13	18

[a] molecule not included into the AD of the model

[b](Ando, Fujimoto, & Morisaki, 1991)

Table 2 – Performances of the previous and new QSPR models

	constitutional	topological	both	previous model[a]
	Eq. 2	Eq. 3	Eq. 4	Eq. 5
R^2	0.84	0.78	0.85	0.90
MAE_{TR} (%)	17	18	16	12
Q^2_{LOO}	0.79	0.71	0.76	0.86
Q^2_{10CV}	0.80	0.71	0.76	-
Q^2_{5CV}	0.78	0.71	0.74	-
R^2_{YS}	0.13	0.10	0.13	-
SD_{YS}	0.08	0.07	0.08	-
R^2_{EXT}	0.81	0.46	0.81	0.84
Q^2_{EXT}	0.81	0.42	0.81	-
MAE_{EXT} (%)	18	32	19	18
R^2_{IN}	0.81	0.82	0.87	0.86
Q^2_{IN}	0.81	0.81	0.87	-
MAE_{IN} (%)	18	15	13	17

[a](Fayet, Rotureau, Joubert, & Adamo, 2011)

Figure 1 – Distributions of the experimental values in the training and validation sets.

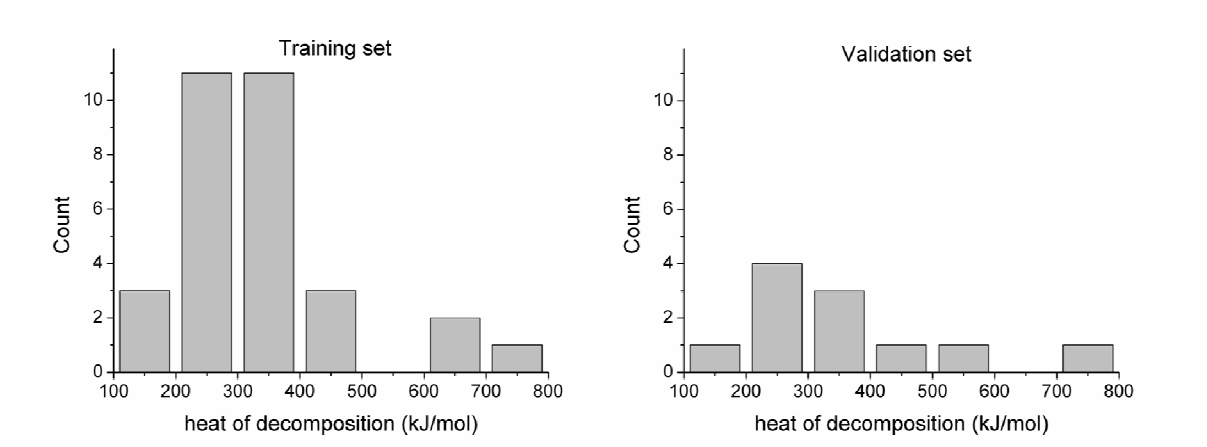

Figure 2 – Experimental vs. predicted heats of decomposition (in kJ/mol) of nitroaromatic compounds based on Eq. 2.

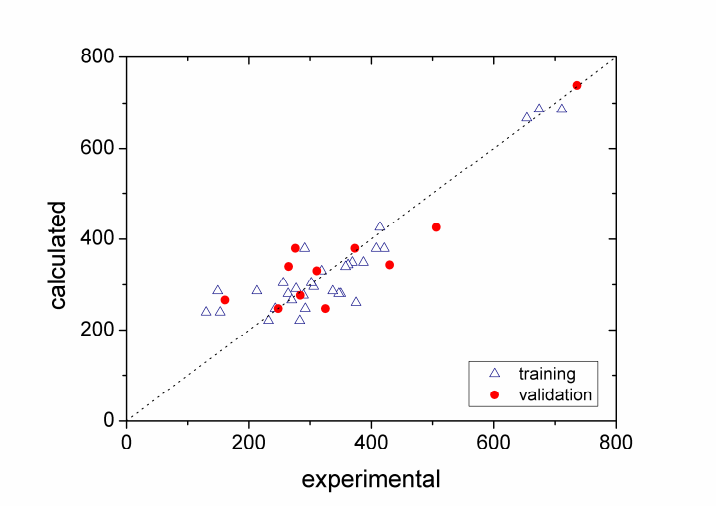

Figure 3 – Correlation of the models issued from Y-randomisation (R^2_{random}) vs. level of randomisation, as estimated by the correlation between the randomised and experimental values ($R^2(Y_{random}/Y_{exp})$).